

# Annotation Guidelines version 1.0: Disambiguating modals in the Vaccination Debate

Liza King

March 2018

## Contents

<b>1</b>	<b>Modality: An Introduction</b>	<b>2</b>
<b>2</b>	<b>Modality Annotation Schema</b>	<b>3</b>
2.1	The Corpus . . . . .	3
2.2	Markables . . . . .	3
2.2.1	Modal . . . . .	3
2.2.2	Subject . . . . .	3
2.2.3	Modalised . . . . .	3
2.2.4	Concession and conditional . . . . .	4
2.3	Links . . . . .	4
2.4	Scope . . . . .	4
<b>3</b>	<b>Instructions</b>	<b>4</b>
<b>4</b>	<b>Examples</b>	<b>5</b>

If you use this guidelines or want to refer to them, please cite the paper:

- Liza King and Roser Morante (2020) Must Children be Vaccinated or not? Annotating Modal Verbs in the Vaccination Debate. Proceedings of LREC 2020, Marseille, France.

# 1 Modality: An Introduction

These guidelines form part of the methodology of research into disambiguation of modal auxiliaries in the domain of the Vaccination Debate. The motivation behind studying modals is to investigate the way in which people qualify their assertions, exemplified in the following sentence pair:

1. Vaccination **COULD** have caused his autism.
2. Vaccination **MUST** have caused his autism.

Both sentences qualify the author’s perspective on the possibility of vaccinations causing autism (an epistemological stance), however, the former sentence lacks the certainty of the latter. Such unambiguous sentence pairs are not too common in user-generated text however, and senses can frequently overlap or be obscured. The aims of this annotation task are then to analyse and resolve these ambiguous cases in a systematic way.

Generally, modality can be defined as ‘a category of linguistic meaning having to do with the expression of possibility and necessity. A modalised sentence locates an underlying or preajacent proposition in the space of possibilities’ [Von Fintel, 2006]. In English, modality can be expressed through modal auxiliaries (*should, might, must*), semimodal verbs (*has to, ought to*), adverbs (*perhaps*), and nouns (*possibility*), among others. This research focuses on modal auxiliaries and the modal senses as described in Ruppenhofer and Rehbein [2012] (Table 1).

	can/could	may/might	must	ought	shall/should
<b>epistemic</b>	+	+	+	+	+
<b>deontic</b>	+	+	+	+	+
<b>dynamic</b>	+	-	-	-	-
<b>optative</b>	-	+	-	-	-
<b>concessive</b>	-	+	-	-	-
<b>conditional</b>	-	-	-	-	+

Table 1: Modal senses classification schema

The senses of Table 1 can be exemplified as follows:

3. He got autism straight after he was vaccinated. It **MUST** be the vaccine.
  - (a) Epistemic: considering the evidence (correlation of onset of autism and vaccination), we can infer that vaccination caused his autism. (Note that this is epistemic modality regardless of logical fallacies.)
4. Infants **MUST** receive the MMR vaccine between 8-12 months.
  - (a) Deontic: implicit agents have the obligation to vaccinate their children between 8-12 months of age.
5. Not all people **CAN** get vaccinated.
  - (a) Dynamic: some people are not able to get vaccinated. In this case it is due to immunocompromisation or some other physical attribute - not because of permission (that would be deontic).
6. Long **MAY** she live.
  - (a) Optative: the author expresses a wish that she lives a long life.
7. Harmful as they **MAY** be, it is still better to vaccinate than not.
  - (a) Concessive: similar to epistemic modals, but here the author considers vaccination to be better than the alternative, and that it is a given rather than a possibility.
8. **SHOULD** you refuse to vaccinate your child, they will not be permitted to enter public school.
  - (a) Conditional: the penalty of refusing to vaccinate your child is that they are refused entry to public school.

The above sentences have the following components: subject, modal auxiliary, and a portion of text that is modalised. These are to form part of the annotation task to provide context. Returning to the first example sentence, *Vaccination* is the subject; *could* is the epistemic modal, and *have caused autism* is the modalised. The auxiliary verb *have* is included in the modalised for purposes of simplicity and so that the entire sentence is annotated. Further detail and examples are provided in the following sections.

## 2 Modality Annotation Schema

### 2.1 The Corpus

The corpus contains 50 documents, retrieved from the Vaccination Corpus, compiled by the Computational Lexicology and Terminology Lab (CLTL) at VU Amsterdam<sup>1</sup>. It contains news articles, blog posts, editorials, governmental reports and science articles about the vaccination debate scraped from the web.

### 2.2 Markables

#### 2.2.1 Modal

The eight modal auxiliaries of Table 1 form the class MODAL, and their corresponding senses are the allowable attributes. The table should provide a useful check when classifying the sense of the encountered modal auxiliary (e.g. when the modal is SHOULD, it is evident that the senses can only be either epistemic, deontic or conditional). Only sentences containing modal auxiliaries are to be annotated, thus it is best to first annotate these with their senses, then follow with annotating the other components subject and modalised, and lastly provide the appropriate relationship links. More examples of the modal senses are in the Example subsection below.

When there are two modals and modalised per sentence that share a subject, link the subject to both modals (unless there is a pronoun that can be linked in the case of the second - then mark the pronoun as the second subject). Do not annotate modals that are part of the ‘about me’ section (such as ‘you can follow me on Twitter’) or (sub)headings.

According to Ruppenhofer and Rehbein [2012], some modal senses cause more disagreement between annotators than others. When the annotator is unsure of the correct sense, they are to use the ‘ambiguous’ modal attribute. This attribute, however, is not to be used too liberally as the aim of this task is not to have perfect inter annotator agreement (IAA), but rather to investigate the type of cases in which disagreement between senses occurs. In practice, if one sense is more likely than another, it is then preferable to mark that sense rather than leave it undecided or ambiguous. In the appendix of these guidelines are example senses from Ruppenhofer and Rehbein [2012] that can aid this decision process.

#### 2.2.2 Subject

SUBJECT is the class that includes the agent in transitive constructions in the active voice, the object/patient when in the passive voice, and the impersonal subject in intransitive stative constructions. The SUBJECT does not have to be animate or human, as in the first two example sentences of this document (where the subject was ‘vaccination’). SUBJECT does not have any attributes to annotate.

The subject noun phrase is to be annotated in full. Sometimes the subject will be far away from the modal and modalised: ‘If not, the baby misses out on important protection gained from it’s mother’s immunity, and it *may* be more susceptible to illness’. Here it is preferable to mark the first subject NP as the subject, rather than the second pronoun *it* as it is more informative. If these were separate sentences, however, the pronoun would be marked as the subject.

#### 2.2.3 Modalised

As the name suggests, the modalised is the span of text (at the clausal level) that is being qualified by the modal auxiliary. It generally has the same span as a predicate, without the modal auxiliary, and can thus contain prepositional phrases and object-position noun phrases as well. Sometimes

<sup>1</sup><https://github.com/ChantalvanSon/VaccinationCorpus>

in user-generated text, sentences are incomplete, or do not conform to the usual structure of NP + predicate, such as the following: ‘It must.’ Here there is no portion of text that is being modalised, and the class is therefore left out, and only the subject and modal are annotated.

### 2.2.4 Concession and conditional

These two markables are the only cases where there are four spans to be annotated instead of the usual three. The naming convention of these two markables are not to be confused with the semantic meaning of ‘concession’ or ‘condition’: these classes are often not the condition or concession in the sentence, but are rather the complement to the modal of that type. To illustrate: ‘while vaccination *may* provide immunity, everyone has a right to decide not to vaccinate.’ The second clause is marked as the class `CONDITION` because it the complement to the conditional modal, but semantically, the first clause is the condition.

When there is an additional modal in the concession or condition span, it is preferable to mark them separately (i.e. mark each modal with its subject and modalised as normal and not as part of a condition or concession). To use the previous example with two modals: ‘while vaccination *may* provide immunity, everyone *should* have a right to decide not to vaccinate’. Both clauses have modals and will therefore be marked separately.

## 2.3 Links

The modal is the core of the annotated clause or sentence, and all other annotated components link to it. There are therefore four possible links:

- Subject -> modal
- Modalised -> modal
- Concession -> modal
- Condition -> modal

## 2.4 Scope

The annotations are limited to the clause or simple sentence level. For most of the modal senses this ensures simplicity, with the exception of concessive and conditional modals. These require a main clause and a dependent clause as the modal forms part of one of the clauses (as seen in sentences 7 and 8 in the introduction section). Elsewhere, only the clause that contains the modal is to be annotated, as this is the target of the current research. Anaphora resolution is out of the scope of the current research, and consequently, the pronouns within the clause that contains the modal will be annotated as the subject or the modalised if there is no available resolution within the sentence, even if it results in ambiguity (see sentence 3 in the Introduction). There is no separate class for negation. When the negation forms part of the subject, it is to be included in the subject, and the same applies for when it forms part of the modalised.

When the modalised sentence is in the form of a question, it is to be marked as one would a statement. The constituent order (modal, then subject, then modalised) will differentiate it from a statement. E.g. ‘Can measles vaccines work to protect against measles? Absolutely, they can...’

## 3 Instructions

The first phase of the annotation task is to annotate all of the modals in the text and label their senses according to Table 1. Following this, the annotators are to mark the modalised, as well as the subject. If the modalised has separate spans, the multiple span function of eHOST is to be used to join them as one annotation. If the modalised clause is incomplete (i.e. no verb, or no event), the modal is still to be annotated and the modalised is left out. All other text that does not form part of the modalised sentence or clause is to be left unannotated.

## 4 Examples

The following six examples are correctly annotated versions from the Introduction section above. Following these are others for further illustration.

9. He got autism straight after he was vaccinated.  $\underline{\text{It}}_{(\text{subj})}$   $\text{MUST}_{(\text{epistemic})}$  *be the vaccine*<sub>(modalised)</sub>.

Anaphora resolution is out of the scope for this task, thus the ‘empty’ subject *it* is annotated and the sentence prior is not annotated.

10. Infants<sub>(subj)</sub>  $\text{MUST}_{(\text{deontic})}$  *receive the MMR vaccine between 8-12 months*<sub>(modalised)</sub>.

11. Not all people<sub>(subj)</sub>  $\text{CAN}_{(\text{dynamic})}$  *get vaccinated*<sub>(modalised)</sub>.

Negation does not have its own annotation class, and thus needs to be included with the element that it negates. In the above case it is within the subject NP ‘not all people’. Similarly, if the negation modifies the modalised, it is to be annotated as part of that as well (e.g. ‘I  $\text{CAN}$  *never get vaccinated*’).

12. *Long*<sub>(modalised-)</sub>  $\text{MAY}_{(\text{optative})}$  she<sub>(subj)</sub> *live*<sub>(-modalised)</sub>.

*Live long* is the modalised in the above sentence, but its span is not contiguous. The hyphen is used to illustrate this here, but it is not necessary to use one during the task as eHOST has a function to mark multiple spans as one annotation.

13. *Harmful as*<sub>(modalised-)</sub> they<sub>(subj)</sub>  $\text{MAY}_{(\text{concessive})}$  *be*<sub>(-modalised)</sub>, [it is still better to vaccinate than not]<sub>(concession)</sub>.

Once more, the above sentence has an event that is not contiguous. ‘As’ is to be included in the modalised in order to annotate the whole clause and the second clause is annotated as the concession.

14.  $\text{SHOULD}_{(\text{conditional})}$  you<sub>(subj)</sub> *refuse to vaccinate your child*<sub>(modalised)</sub>, [they will not be permitted to enter public school]<sub>(condition)</sub>.

As with sentence 13, the second clause is annotated, this time as the conditional complement to the modal ‘should’.

## References

Josef Ruppenhofer and Ines Rehbein. Yes we can!? annotating the senses of english modal verbs. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)*, pages 24–26, 2012.

Kai Von Fintel. Modality and language. 2006.

## Appendix

<b>Modal</b>	<b>Sense</b>	<b>Example</b>
Can	Dynamic	These animals can be aggressive The wind can still get in.
	Deontic	You can come in now.
	Epistemic	He says it was Jill but it can't have been. That can't be her - she's twenty-five, that woman is at least 45.
Could	Dynamic	My father could run really fast when he was younger.
	Deontic	I knocked and she said I could come in.
	Epistemic	The sportsman hinted he could be in the lineup.
May	Epistemic	It may hurt a little. He may be home now as it's past 7 o'clock.
	Deontic	Yes, you may come in.
	Concessive	He may be a professor, but he is still a fool.
	Optative	May you live to 100.
Might	Epistemic	He might catch the train on time.
Must	Epistemic	Someone must be home because the lights are on.
	Deontic	The door must remain shut at all times. You must be on time tomorrow.
Should	Epistemic	Adequate sunblock should prevent you from sunburn.
	Deontic	You should dress appropriately.
	Conditional	Should you see him, please tell him to call me. Should you hear the alarm, form an orderly queue.
Ought	Epistemic	He ought to be home as it is past 7 o'clock.
	Deontic	She ought to apologise.

Table 2: Examples of modals and their senses (adapted from [Ruppenhofer and Rehbein \[2012\]](#))

<b>Sense</b>	<b>Explanation</b>
Dynamic	Concerning ability or potential for involvement in events or behaviour
Deontic	Concerns giving permission
Epistemic	Concerns the possibility for the speaker to come to certain conclusions
Concessive	Compatible with epistemic sense (see May), but used when the speaker thinks the proposition is true, and not just a possibility
Optative	Used to express a wish
Conditional	Contains a conditional component, sometimes with subject-auxiliary inversion

Table 3: Senses and explanation (adapted from [Ruppenhofer and Rehbein \[2012\]](#))